

# 基于隐因子模型的 P2P 借贷推荐算法

李友<sup>1</sup>, 马晔赞<sup>2</sup>, 陈杨阳<sup>3</sup>

(1. 厦门大学信息科学与技术学院, 福建 厦门 361005; 2. 北京大学汇丰商学院, 广东 深圳 518000; 3. 内蒙古大学计算机学院, 呼和浩特 010000)

摘要: P2P 网络借贷的兴起使得网络借贷平台上投资用户数目以及贷款项目数目迅速激增, 同时带来海量的借款人信息以及贷款信息, 投资用户在众多贷款项目中找到自己感兴趣的贷款项目变得愈加困难。本文在研究推荐系统相关算法的基础上, 详尽的实验了基于隐因子模型的推荐算法, 在挖掘用户感兴趣的贷款的准确率。为了说明算法的有效性, 本文做了相关推荐算法的多组对比实验。对比实验结果表明, 基于隐因子模型的推荐算法能够更有效地预测投资用户的投资偏好, 并且对推荐系统应用到 P2P 借贷平台具有借鉴意义。

关键词: P2P 借贷; 推荐系统; 隐因子模型  
DOI: 10.16640/j.cnki.37-1222/t.2016.04.233

## 0 引言

伴随信息技术的迅速发展, 互联网金融在近年来迅速兴起。其中 P2P 网络借贷凭借低成本, 快速便捷等特点正成为人们投资理财的新选择。网络借贷平台是新兴的金融市场, 通过互联网通道产生的海量交易数据促使了许多相关学术研究。关于网络借贷的研究大致可以分为对投资者出借意愿的研究<sup>[1]</sup>、对贷款项目评估的研究<sup>[2]</sup>以及对网络借贷中社交网络的研究<sup>[3]</sup>。

本文的研究内容是使用推荐算法, 分析用户的相关信息, 从正在招标的贷款项目中挖掘出用户可能感兴趣、可能投资的项目, 并推荐

给相应的投资用户, 实现提高网络借贷平台上借贷双方的匹配效率的目的。本文的创新工作: (1) 从信息检索和机器学习方法的角度考虑, 挖掘并推荐给出借人感兴趣的项目, 实现提高资本需求双方匹配效率进而提高平台运营效率的研究还是从未有过的; (2) 推荐系统目前主要应用于电子商务与社交网络领域, 将其运用到网络借贷平台上的情况也不曾有; (3) 使用具有高可扩展性的机器学习方法 -- 矩阵分解技术<sup>[4]</sup>, 将推荐系统<sup>[5]</sup>应用到网络借贷平台。

## 1 相关知识

本部分将介绍论文研究所要用到的相关知识以及本文实验数据

下降时间最小值, 占空比最大值, 占空比最小值, 脉冲个数(齿圈对数)。在测量过程中还会时时显示波形以使用户观测。伺服电机带动工件的旋转速度用户可以在系统里自行设置, 默认为 60rpm。

数据管理功能包含数据存储, 数据查询、追溯, 数据报表, SPC 分析等功能, 且可以导出至 excel 文件, 通过以太网或 U 口导出。下图 4 为数据查询部分界面。



图 4

用户设置功能包含用户的添加、删除、密码修改, 电机转速的设置, 数据判定范围的设置, 采样率的设置等。

校准功能为用户任意设定一定幅值、频率的波形, 通过模拟量输出板卡输出, 之后再由本系统的模拟量采集卡采集输入, 进行判定, 如误差超过设置范围, 则报警, 通知管理人员进行分析维护。

另外, 本系统还包含数据滤波模块。其中, 滤波算法中应用了 LabVIEW 里自带的 FFT 模块。FFT 是离散傅立叶变换的快速算法, 是把一个时域信号转换到频率里进行分析。在系统测试时, 系统的噪声, 机械的振动等都会对测试数据带来干扰, 本系统对无效的频率抛弃, 而后对有效的频率数据进行还原。部分滤波程序见下图 5。

滤波对比见下图 6, 其中红色为带干扰波形, 蓝色为还原后波形, 如图所示, 滤波后的数据可以去掉干扰。

## 4 结论

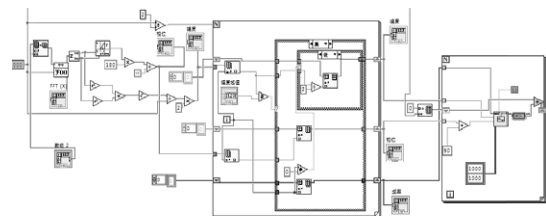


图 5

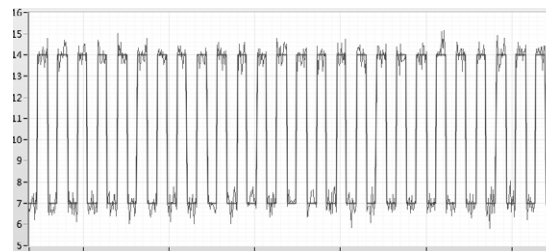


图 6

综上所述, 本文所述的测试系统, 提高了产品质量, 简化了测试过程, 工作稳定, 节省成本, 可以满足工业生产线上大批量的在线检测要求。

## 参考文献:

[1] 林游, 张俊杰, 易凡. 霍尔传感器信号采集与显示系统设计 [J]. 现代电子技术, 2009, 32(04): 199-202.  
[2] 陈锡辉, 张银鸿. LabVIEW 8.20 程序设计从入门到精通 [M]. 北京: 清华大学出版社, 2007.  
[3] 钱同惠. 数字信号处理 [M]. 北京: 机械工业出版社, 2005.

集。

1.1 协同过滤推荐算法

协同过滤推荐是基于这样的假设：如果用户之间的历史行为与偏好是相似的，那么我们可以向他们推荐他们之间相互感兴趣但未发掘的内容。

其方法是将原有的信息构建成一个用户评分矩阵 R，行向量代表每一个用户，列向量代表每一个项目，其中的每一个值代表特定用户对特定项目的评分（喜好程度）。目前协同过滤算法主要有基于内存的推荐和基于模型的推荐。矩阵分解即是属于基于模型 [5]。

1.2 本文数据介绍

由于我国网络借贷平台数据未被公开，所以无法获得，故本文数据采用的是美国的网络借贷平台 Prosper.com 上的数据。本文实验只需用到 Bids 与 Loans 表。Bids 表保存了投资者的出借记录，Loans 表记录了借款项目的基本信息和状态。

在网络借贷中，并没有投资者对贷款的评分。但从投资用户对贷款项目的投资金额来看，可以预见，越多的出借数额代表投资用户对该贷款项目越有兴趣，评分也越高。即将投资金额离散为 10 个等级，从 1 至 10 表示金额越大则级别越高，级别越高则表明越感兴趣，我们称它为“投资等级”。

2 基于隐因子模型的推荐算法

矩阵分解模型是当前基于模型的协同过滤算法中应用最流行的一种 [6]，并且获得了较理想的推荐效果。Yehuda Koren 在 Netflix 比赛中使用称作 LFM (latent factorization model) 的矩阵分解方法，推荐效果得到显著提升。

2.1 隐因子模型算法实现

对于 M\*N 的评分矩阵，传统的是使用 SVD 进行分解，而实际上这种直接计算特征值的传统方法的计算复杂度很高，在大规模的矩阵上基本不可行，而实际中系统动辄就是上千万的用户和上百万的物品，所以这种方法无法使用。因此，在实际应用中我们使用 LFM 方法进行矩阵分解求解。LFM 方法的形式化描述如下所示：

$$\min U, V \sum_{i=1}^M \sum_{j=1}^N \|R_{ij} - U_i^T V_j\| + \lambda (\|U_i\|^2 + \|V_j\|^2) \quad (1)$$

其中  $U_i$  表示的是一个 f 维的用户因子向量， $V_j$  代表的是一个 f 维的项目因子向量； $U_i^T V_j$  表示用户 i 对贷款 j 的估计投资等级。

求解目标函数 (1)，通常有两种方法，一种是交叉最小二乘法，一种是随机梯度下降法。由于交叉最小二乘法计算复杂且运算速度慢，所以使用随机梯度下降法。

2.2 实验介绍

(1) 数据预处理。从 Bids 表中导出每个投资者的投资记录，Loans 表中导出每个借款成功的贷款项目记录，并按投标日期排序。由于 Prosper.com 公开的数据有限，为了保证信息的及时性，我们采用最新的 2009 年至 2010 年的投资记录进行实验并评估推荐效率。

(2) 结果分析。经过隐因子分解之后计算内积得到原矩阵的逼近矩阵，得出了每个用户对借款的投资等级预测值。在评估我们的实验算法时，我们使用评估推荐效率常用的算法：采纳率、精确率、召回率和 F 值来表示。实验结果发现，随着推荐项目的增多，各项评估指标也增大。当对每个投资用户推荐 10 个贷款时，最终有 86% 的用户至少投资了 10 个贷款中的一项，并且精确率、召回率和 F 值各项指标也分别达到 24%、43% 和 30.8%。

基于内存的推荐也是协同过滤算法家族使用最为广泛的推荐算法之一。基于内存的推荐的投资等级预测公式如下：

$$p_{u,j} = \bar{v}_u + \alpha \sum_{i=1}^n \omega_{u,i} (v_{i,j} - \bar{v}_i) \quad (2)$$

在式中， $\bar{v}_u$  表示用户 u 的平均投资等级； $\omega_{u,i}$  表示用户 u 和 i 之间的相关性，在实验中我们选取皮尔森相关性系数来计算相关性； $\alpha$  是归一化因子；n 表示参照用户数，对 n 的选取，常用的有两种方式，一种是指定相关度阈值，一种是选择固定的 K 个邻居数。

为说明基于隐因子模型的算法比基于内存的推荐算法相对有效，我们使用同样的训练和测试数据集，和使用指定密度阈值和邻居数 K 的基于内存的协同过滤推荐算法在最优时的表现对比（经过多次实验得出  $\alpha=0.1$  和  $K=15$ ）。实验结果表明，在推荐 10 个贷款项时，LFM 的推荐准确性要优于基于内存的推荐算法超过 10%。

3 基于借贷行为差异的改进算法

在现实生活中，投资用户的出借行为是存在差异的。某些投资者的闲散资金比较充裕，他们的平均投资额超过普遍的用户；同时，贷款之间也存在差异。如有的借款被看好，投资用户愿意出借更多的资金，而使得贷款的平均出借额更高。因此，借款人的投资行为以及贷款项目的优劣都会对出借额的多少产生影响。

3.1 模型改进

现将借贷行为差异考虑到 (1) 中的模型内。若全体投资用户对一个项目的平均出借等级为 A，用户 i 对一个项目的平均出借等级比 A 高出  $B_i$  ( $B_i$  大于 0 即表示 i 的投资等级高于平均值，小于 0 则表示低于平均值)；同理，若全体贷款的每一个出借人的平均出借等级为 C，贷款 j 目前的平均出借等级比 C 高出  $L_j$  ( $L_j$  大于 0 即表示 j 的出借等级高于平均水平，小于 0 则低于平均水平)，这里的  $B_i$  与  $L_j$  即为出借人的出借偏好和贷款项目的好坏差异。则用户 i 对贷款 j 的估出投资等级为  $\hat{R}_{ij}$  的表达式修改为下式：

$$\hat{R}_{ij} = U_i^T V_j + L_j + B_i \quad (3)$$

同样优化公式将修改为下式：

$$\min U, V \sum_{i=1}^M \sum_{j=1}^N \|R_{ij} - \hat{R}_{ij}\| + \lambda (\|U_i\|^2 + \|V_j\|^2 + \|B_i\|^2) \quad (4)$$

3.2 结果分析

将改进后的模型重新做预测，可以发现精确率、召回率、F 值都有部分提升。当向每个用户推荐 10 个贷款时，最终有 89% 的用户至少投资了其中一项，可以发现采纳率在改进后提高相对明显，对于推荐特定 K 个项目，改进后的采纳率平均提高 4%。

4 结束语

对于网络借贷平台，如果能够使用有效的推荐系统，向投资用户推荐他们感兴趣的贷款项目，不仅可以减少用户在投资决策时的信息筛选，提升用户体验，还能提高借款效率而提高运营效率。通过 Prosper 网站真实数据实验，结果表明，基于隐因子模型的推荐算法可以有效的利用用户的历史投资记录，找出用户感兴趣的贷款项目；并且当考虑不同用户与贷款项目的个体偏见时，可以更好地提升推荐效率。

参考文献：

[1] 宋文, 韩丽川. P2P 网络借贷中投资者出借意愿影响因素分析 [J]. 西南民族大学学报: 自然科学版, 2013, 39(05): 795-799.  
 [2] Li S, Qiu J, Lin Z, et al. Do borrowers make homogeneous decisions in online P2P lending market? An empirical study of PPDai in China[C]// Service Systems and Service Management (ICSSSM), 2011 8th International Conference on. IEEE, 2011: 1-6.  
 [3] Berger S C, Gleisner F. Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending[J]. BuR - Business Research, 2010, 2(01): 39-65.  
 [4] 程明松, 刘勺连. 一种实用快速非负矩阵分解算法 [J]. 大连理工大学学报, 2013, 53(01): 151-156.  
 [5] J. Bobadilla, F. Ortega, A. Hernando, A. Gutierrez. Recommender systems survey. Knowledge-Based Systems, 2013, 46(01): 109-132.

作者简介：李友 (1990-)，男，江西赣州人，硕士研究生，主研领域：数据挖掘。